GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

QM nucleic - nucleic search, using sw model

Run on: March 14, 2002, 14:28:55 ; Search time 1725.6 Seconds
(without alignments)
411.097 Million cell updates/sec

Title:
Perfect score: 43
Sequence: 1 gattgatcaaacaaatctact.....tgctaattttttttaaaa 43

Scoring table: IDENTITY_NUC
Gap 10.0, Gapext 1.0

Searched: 147x148 seqs, 92185997755 residues

Total number of hits satisfying chosen parameters: 294128

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_pr:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_ss:*
12: gb_sts:*
13: gb_vi:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_pr:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_pr:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | | Description |
|---|---|---|---|---|---|---|
| 1 | 23.6 | 54.9 | 1.0000 | 2 | HSS171M_2 | Continuation ( ) of |
| 2 | 23.6 | 54.9 | 4[illegible] | 9 | [illegible] | AL163302 Homo sapi |
| 3 | 23.4 | 54.4 | 2.6259 | 2 | AL583892 | AL583892 Mus muscu |
| c 4 | 23 | 53.5 | 3[illegible] | 1 | [illegible] | AL109989 Streptomy |
| 5 | 22.8 | 53.0 | 4[illegible] | 2 | [illegible] | A[illegible] |
| 6 | 22.8 | 53.0 | 1.2759 | 2 | AL[illegible] | AL[illegible] Hom |
| 7 | 22.8 | [illegible] | [illegible] | 2 | [illegible] | A[illegible] Hom |
| c 8 | 22.4 | 52.1 | 4771 | 1 | [illegible]44 | A[illegible] |
| c 9 | 22.4 | 52.1 | 4.266e-08 | 9 | AC041[illegible] | AC041[illegible] Hom |
| 10 | 22.4 | 52.1 | 4.994e-08 | 2 | AC[illegible] | AC[illegible] Hom |
| c 11 | 22.4 | 52.1 | 5.644e-08 | 2 | AF[illegible] | AF[illegible] Mus |
| c 12 | 22.4 | 52.1 | 5.997e-08 | 2 | AC[illegible] | AC[illegible] Mus |
| c 13 | 22.2 | 51.6 | 1.599e-07 | 9 | AE[illegible] | AE[illegible] Hom |
| 14 | 22.2 | 51.6 | 1.908e-07 | 2 | AC[illegible] | AC[illegible] Hom |
| c 15 | 22.2 | 51.6 | 1.908e-07 | 2 | AC[illegible] | AC[illegible] Hom |
| 16 | 22.2 | 51.6 | 1.908e-07 | 2 | AC[illegible] | AC[illegible] Hom |
| c 17 | 22.2 | 51.6 | 1.908e-07 | 2 | AC[illegible] | AC[illegible] Hom |
| c 18 | 22 | 51.2 | 3.564e-07 | 2 | AC[illegible] | AC[illegible] Hom |
| c 19 | 22 | 51.2 | 3.564e-07 | 2 | AC[illegible] | AC[illegible] Hom |
| 20 | 21.8 | 50.7 | 6[illegible] | 1 | [illegible] | AE[illegible] Str |
| c 21 | 21.8 | 50.7 | 67948 | 1 | AE007[illegible] | AE[illegible] Str |
| c 22 | 21.8 | 50.7 | 6.659e-07 | 1 | AE[illegible] | AE[illegible] Hom |
| 23 | 21.8 | 50.7 | 6.659e-07 | 2 | AC[illegible] | AC[illegible] Hom |
| 24 | 21.6 | 50.2 | 2.042e-06 | 2 | AE[illegible] | AE[illegible] |
| c 25 | 21.6 | 50.2 | 419 | 6 | S[illegible] | M[illegible] |
| 26 | 21.4 | 49.8 | 5[illegible] | 10 | AE[illegible] | AE[illegible] M |
| 27 | 21.4 | 49[illegible] | 4[illegible] | 2 | AE[illegible] | A[illegible] Ps |
| c 28 | 21.2 | 49[illegible] | 1.03e-06 | 2 | AE[illegible] | AE[illegible] Es |
| c 29 | 21.2 | 49.3 | 1.079e-06 | 1 | AE[illegible] | AE[illegible] Es |
| 30 | 21.2 | 49[illegible] | 1e-[illegible] | 2 | AC[illegible] | AC[illegible] Dr |
| 31 | 21.2 | 49.3 | 22527 | 2 | AC023[illegible] | AC[illegible] Dr |
| 32 | 21.2 | 49.3 | 33[illegible] | 2 | AC[illegible] | AC[illegible] Dr |
| c 33 | 21.2 | 49.3 | 11[illegible] | 1 | [illegible] | [illegible] |
| c 34 | 21.2 | 49.3 | 1[illegible] | 1 | AE001[illegible] | AE001048 Chlamydi |
| c 35 | 21.2 | 49.3 | 1[illegible] | 2 | AC[illegible] | AC[illegible] M |
| 36 | 21.2 | 49.3 | 1683[illegible] | 2 | AC[illegible] | AC[illegible] Mus |
| 37 | 21.2 | 49.3 | 17[illegible] | 2 | AE[illegible] | AE[illegible] |
| c 38 | 2[illegible] | 48.8 | [illegible] | 1 | AX[illegible] | AX[illegible] |
| c 39 | 2[illegible] | 48.8 | [illegible] | 1 | AX[illegible] | AX[illegible] |
| 40 | 2[illegible] | 48.8 | 6[illegible] | 2 | AC[illegible] | AC[illegible] |
| 41 | 2[illegible] | 48.8 | 7801[illegible] | 2 | AC[illegible] | AC[illegible] |
| c 42 | 2[illegible] | 48.8 | 8[illegible] | 1 | AE[illegible] | AE[illegible] |
| 43 | 2[illegible] | 48.8 | 8[illegible] | 2 | AC[illegible] | AC[illegible] |
| c 44 | 2[illegible] | 48.8 | 8[illegible] | 2 | AC[illegible] | AC[illegible] |
| 45 | 2[illegible] | 48.8 | 8[illegible] | 2 | AC[illegible] | AC[illegible] |

ALIGNMENTS

RESULT 1
HSS171M_2
WPCOMMENT
Sequence split into 4 fragments [100 S[illegible]171M [illegible] 2[illegible]
Fragment Name Begin End
HSS171M_0 1 [illegible]
HSS171M_1 1[illegible] [illegible]
HSS171M_2 2[illegible] [illegible]
HSS171M_3 3[illegible] 4[illegible]
Continuation (3 of 4) of HSS17[illegible] from base [illegible] 2[illegible] Homo sapiens [illegible] chromosome

Query Match 54.9%; Score [illegible]; [illegible] mismatches; [illegible]
Best Local Similarity 70.4%; Pred. No. [illegible]
Matches 2[illegible]; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible]

Qy 6 gtcaaacaaatctactc[illegible] 43
|| [illegible]
Db 1741 GTAAATCAAAT-[illegible] 1[illegible]

```
RESULT  2
HS210102
LOCUS       HS210102          [illegible] bp    DNA             PRI       24 MAY 2000
DEFINITION  Homo sapiens chromosome 21 segment HS210102.
ACCESSION   AL[illegible]
VERSION     AL[illegible]
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 46000)
  AUTHORS   Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
            Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
            Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
            Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
            Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
            Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
            Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,
            Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
            Hornischer,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A.,
            Reichelt,J., Kauer,G., Blocker,H., Ramser,J., Beck,A., Klages,S.,
            Hennebold,S., Maier,E., Wimmer,H., Rudolph,A., Boxem,K.,
            Cardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
            Yaspo,M.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (   MAY 2000) The Chromosome 21 Mapping and Sequencing
            Consortium * RIKEN Genomic Sciences Center, Human Genome Research
            Group * Institute of Molecular Biotechnology, Genome Analysis * *
            Keio University School of Medicine, Dept. of Molecular Biology *
            GBF, Dept. of Genome Analysis * Max Planck Institute for Molecular
            Genetics (addresses see below)
COMMENT     The Chromosome 21 Mapping and Sequencing Consortium consists of
            * RIKEN Genomic Sciences Center, Human Genome Research Group, *
            Saitama-ken 230-0045, Japan,
            * e-mail: sakaki@gsc.riken.go.jp
            * URL: http://[illegible]
            and
            * Institute of Molecular Biotechnology, Genome Analysis, *
            Beutenbergstrasse 11, D-07745 Jena, Germany,
            * e-mail: qsc@submit.genome.imb-jena.de
            * URL: http://qsc.imb-jena.de/
            and
            * Keio University School of Medicine, Dept. of Molecular Biology, *
            Tokyo 160-8582, Japan,
            * e-mail: shimizu@sc.kusm.med.keio.ac.jp
            * URL: http://genome.imb.med.keio.ac.jp/
            and
            * GBF, Dept. of Genome Analysis,
            * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e-mail:
            info-genome@gbf.de
            * URL: http://www.gbf.de/
            and
            * Max Planck Institute for Molecular Genetics,
            * Ihnestrasse 73, D-14195 Berlin, Germany,
            * e-mail: info-chr21@molgen.mpg.de
            * URL: http://[illegible]
FEATURES             Location/Qualifiers
     source          1..[illegible]
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.3"
     source          [illegible]
                     /note="Accession No. AL[illegible]"
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.3"
                     /clone="PAC [illegible] partial"
                     /clone_lib="RPCI1,8-9 PAC library"
     repeat_region   2044..2254
                     /note="MER1B"
                     /rpt_family="LTR/MaLR"
                     /rpt_type=DISPERSED
     repeat_region   3576..3912
                     /note="MER11A"
                     /rpt_family="LTR/MaLR"
                     /rpt_type=DISPERSED
     repeat_region   complement(4043..4108)
                     /note="L2M4"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   complement(4116..4440)
                     /note="AluSx"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   complement(4462..4540)
                     /note="L1M4"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   complement(4579..4938)
                     /note="L1M4"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   6423..6466
                     /note="(CA)n"
                     /rpt_family="Simple_repeat"
                     /rpt_type=TANDEM
     repeat_region   7939..[illegible]
                     /note="MER11A"
                     /rpt_family="ERV/MaLR"
                     /rpt_type=DISPERSED
     repeat_region   8[illegible]
                     /note="(CA)n"
                     /rpt_family="Simple_repeat"
                     /rpt_type=TANDEM
     repeat_region   9039..9086
                     /note="(CA)n"
                     /rpt_family="Simple_repeat"
                     /rpt_type=TANDEM
     repeat_region   9681..9975
                     /note="AluSq"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   10626..[illegible]
                     /note="MLT2B"
                     /rpt_family="LTR/Retroviral"
                     /rpt_type=DISPERSED
     repeat_region   complement(10452..10487)
                     /note="L1M"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   10513..10547
                     /note="(A)6"
                     /rpt_family="Simple_repeat"
                     /rpt_type=TANDEM
     repeat_region   10[illegible]
                     /note="(TCCA)3"
                     /rpt_family="Simple_repeat"
                     /rpt_type=TANDEM
     repeat_region   10711..12174
                     /note="L1M"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   12876..13[illegible]
                     /note="L1P"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   complement(12917..13571)
                     /note="L1MA10"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   13565..13[illegible]4
                     /note="L1MA10"
```

```
                     /rpt_family "LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   complement(15672..15818)
                     /note "MER58C"
                     /rpt_family "DNA/MER1_type"
                     /rpt_type DISPERSED
     repeat_region   15*73..16018
                     /note="MER5A"
                     /rpt_family "DNA/MER1_type"
                     /rpt_type=DISPERSED
     repeat_region   complement(16149..16241)
                     /note "MER5A"
                     /rpt_family "DNA/MER1_type"
                     /rpt_type=DISPERSED
     repeat_region   19308..19479
                     /note "MER5A"
                     /rpt_family "DNA/MER1_type"
                     /rpt_type=DISPERSED
     repeat_region   complement(20501..20889)
                     /note "MER58B"
                     /rpt_family "DNA/MER1_type"
                     /rpt_type=DISPERSED
     repeat_region   20927..20968
                     /note "AluSx"
                     /rpt_family "SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   25796..25856
                     /note="L2"
                     /rpt_family "LINE/L2"
                     /rpt_type=DISPERSED
     repeat_region   complement(25939..26049)
                     /note="L2"
                     /rpt_family "LINE/L2"
                     /rpt_type=DISPERSED
     repeat_region   26786..26816
                     /note="MIR"
                     /rpt_family "SINE/MIR"
                     /rpt_type=DISPERSED
     repeat_region   complement(27166..27357)
                     /note "MER61B"
                     /rpt_family "DNA/MER4-group"
                     /rpt_type DISPERSED
     misc_feature    27525..27625
                     /note "Non-B site"
     repeat_region   complement(27605..27979)
                     /note "MLT1B"
                     /rpt_family "LTR/MaLR"
                     /rpt_type DISPERSED
     repeat_region   complement(30048..30285)
                     /note="MER96B"
                     /rpt_family "DNA/MER1_type?"
                     /rpt_type=DISPERSED
     repeat_region   31101..31198
                     /note "(TG)n"
                     /rpt_family "Simple_repeat"
                     /rpt_type TANDEM
     repeat_region   31210..31383

  Query Match             54.9%;  Score 23.6;  DB 9;  Length 340000;
  Best Local Similarity   76.8%;  Prod. No. 26;
  Matches   23;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      1 atcaacgacatctatcaacctcaacgattcatatgtca 43
          ||||||||||||||||||||||||||||||||||||| |||
Db   8734* GTAACGCATCTATCAACCTCAACGATTCATATGTCA 8778*


RESULT   8
AL583892
LOCUS       AL583892    216258 bp    DNA             HTG       26-MAY-2001
DEFINITION  Mus musculus chromosome 15 clone RP23-3267, *** SEQUENCING IN
            PROGRESS ***, 11 unordered pieces.
ACCESSION   AL583892
VERSION     AL583892.4  GI:19115011
KEYWORDS    HTG; HIGH-THROUGHPUT GENOMIC; PHASE1; SEQ_LEN 216258.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
            Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 216258)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (MAY 2001) Sarah Sims, Sanger Centre, Hinxton Hall,
            CB10 1SA, UK, E-mail: ss3@sanger.ac.uk, Tel: 44 1223 494644,
            Fax: 44 1223 494919
COMMENT     ***SEQUENCING IN PROGRESS***  This sequence is placed in
            Contigs 1 and 2.
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: sanger@sanger.ac.uk
            Center project name: IMS_MT
            Submission: 5-JUN-2001
            Assembly program: ARACHNE, version 2
            Sequencing vector: Unknown; 4 clone reads
            Chemistry: type terminator; 4 dye reads
            Consensus quality: 21.4x2 bases > QV 64
            Consensus quality: 2.8x bases > QV 50
            Consensus quality: 2.7x bases > QV 40
            Insert size: 2140*50 sd=*** reads
            Insert size: 2***** 5** reads  ***
            Quality coverage: 1.2x ... bases QV40 or better is quality
            coverage: 0.8x ... bases in *** gaps
            .
            * NOTE: This is a 'working draft' sequence,
            * that may contain errors and may be incomplete.
            * As soon as it is available and finished, this sequence will
            * be promptly updated,
FEATURES             Location/Qualifiers
     source          1..216258
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="15"
                     /clone="RP23-3267"
                     /clone_lib="RPCI-23"
     misc_feature    1..2***
                     /note="assembly_fragment:04
                     fragment_1..2***;
                     clone_end:no
                     vector_shared:"
     misc_feature    1,81*..2****
                     /note="assembly_fragment:02;
                     fragment_1,clone1"
     misc_feature    2*215..4,9**
                     /note="assembly_fragment:...
                     fragment_1,clone2"
     misc_feature    4,492..4*2,8**
                     /note="assembly_fragment:04
                     fragment_1,clone2"
     misc_feature    4*814..,51*,*
                     /note="assembly_fragment:04;
                     fragment_1,clone1"
     misc_feature    *34**..,8*15*
                     /note="assembly_fragment:04
                     fragment_1,clone1"
     misc_feature    *,24*..,*8..,
                     /note="assembly_fragment:04;
                     fragment_1,clone1"
     misc_feature    *8*02..,711**
                     /note="assembly_fragment:...
                     fragment_1,clone1"
     misc_feature    *1,9*1..,1**,2*
                     /note="assembly_fragment:...
                     fragment_1,clone1"
     misc_feature    ,9*1*..,1..,**
```

```
                     /note "assembly fragment:cos05"
                     fragment chain:2"
     misc_feature    13041..13676
                     /note "assembly fragment:cos06"
                     fragment chain:2"
     misc_feature    13777..14606
                     /note "assembly fragment:cos07"
                     fragment chain:2"
     misc_feature    14761..17268
                     /note "assembly fragment:cos08"
                     fragment chain:2"
     misc_feature    17503..18546
                     /note "assembly fragment:cos09"
                     fragment chain:2"
     misc_feature    18947..21627
                     /note "cos10", fragment 2/2"
                     fragment chain:2
                     clone end:17
                     vector side:right"
BASE COUNT    90471 a  52721 c  44061 g  52103 t    1431 others
ORIGIN


  Query Match             54.4%;  Score 23.4;  DB 2;  Length 215200;
  Best Local Similarity   73.2%;  Pred. No. 52;
  Matches   41;  Conservative   0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1 ttttttgtcgcgcgcgtcatgcgcctgccgttgatcgctg 41
           | | | ||| | |||||||| | || ||| | |||| |
Db  157466 gcttacgtcacgcgaagctacgcgcatcacgcttcacgcg 157506


EPSON1    4
SV1L/2
locus       SCJ12          35302 bp    DNA             BCT       24 AUG 1999
DEFINITION  Streptomyces coelicolor cosmid J12.
ACCESSION   AL109989
VERSION     AL109989.1  GI:5777663
KEYWORDS    acetyltransferase; dehydrogenase; DNA binding regulator; integral
            membrane protein; lipoprotein; luxR family; lyase; oxidoreductase;
            methyltransferase; nitrate reductase alpha chain; nitrate reductase
            beta chain; nitrate reductase delta chain; nitrate reductase gamma
            chain; nitrate/nitrite transporter protein; oxidoreductase;
            pyruvate formate lyase activating enzyme; pyruvate phosphate
            dikinase; sigma factor; transposase pseudogene; two component
            response regulator; two component sensor.
SOURCE      Streptomyces coelicolor A3(2).
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 35302)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 32 (2), 343-357 (1999)
  MEDLINE   99268514
REFERENCE   2  (bases 1 to 35302)
  AUTHORS   Murphy,L. and Harris,D.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 35302)
  AUTHORS   Thomson,N.R., Barkhill,J.E., Bartell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (24 AUG 1999) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK; Barrell Sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT     NOTICE:
            Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC and Beowulf Genomics
            Details of S. coelicolor sequencing at The Sanger Centre are
            available on the World Wide Web.
```

```
            URL: http://www.sanger.ac.uk/Projects/S_coelicolor/)
            CDS are named according to the following system of SCJ12.01c, SC (S.
            coelicolor), J12 (cosmid name), 01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous.
            The length in codons is given for each CDS.
            Usually the highest scoring match found by fasta is also given for
            CDS which show significant similarity to other CDS in the database,
            the position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon.
            Gene prediction is based on positional base preference in codons
            using a specially developed Hidden Markov Model (Krogh et al.,
            Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot
            program of Bibb et al., Gene 30:157-66(1984) as implemented at
            http://www.sanger.ac.uk.
            Unusually GTG codons are used as the initiation codon in some CDS.
            Where this occurs we may not have predicted the correct initiation
            codon.  Where possible we choose an initiation codon (atg, gtg, ttg
            or (att)) which is preceded by an upstream ribosome binding site
            sequence (optimally 5-14bp before the initiation codon).  If this
            cannot be identified we choose the most upstream initiation codon.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone.  It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions.
            Cosmid J12 overlaps with cosmid J1 on the AseI-J genomic
            restriction fragment.
FEATURES             Location/Qualifiers
     source          1..35302
                     /organism="Streptomyces coelicolor A3(2)"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
                     /clone="cosmid J12"
     misc_feature    1..183
                     /note="fragment 1/2 of 36793..36925 overlap with cosmid SCJ1"
     gene            complement(1..836)
                     /gene="crt1"
     CDS             complement(1..836)
                     /gene="crtU"
                     /note="SCJ12.01c, probable dehydrogenase, partial CDS,
                     len: >278 aa. Highly similar to Streptomyces griseus
                     TR:P72443 (EMBL: X95596) dehydrogenase (>17 aa), fasta
                     scores: opt: >111, z-score: 1471.5, E(): 0, 74.0% identity in
                     269 aa overlap, and Synechocystis sp. (strain PCC 6803)
                     SW:CRTI_SYNY3 (EMBL: X62574) phytoene dehydrogenase (EC
                     1.3.-.-) (phytoene desaturase) (427 aa), fasta scores:
                     opt: 270, z-score: 256.2, E(): 1.4e-06, 27.6% identity in
                     254 aa overlap."
                     /codon_start=1
                     /transl_table=11
                     /product="putative dehydrogenase CrtU"
                     /protein_id="CAB53413.1"
                     /db_xref="GI:5777664"
                     /translation="MTVRKTAAAKKGKKKKAECIVFFAFGAERFSKGLEPGVAVIEGGI
                     AGLAAAILMALGGAKVTLYEKEAVGGASIKKTTLAPKGAVIMEKGLHAFTKQYYNL
                     RGLLREFIPQFCLSELPMYELKHSGGLTDDGFARVERTPPLSATGEVALSFLPWRDL
                     AAMFRAALTLSVRVPVYEKTAGVIAIAFLEGVKTLFAAHHLAFLVFGKSFFADPR
                     RLSAAELLLMFRIYFLGSSEGLLFDVPVEFFPQALWNFLGDYLRRIGADVRTGFPVRG
                     LLE"
     gene            complement(844..1573)
                     /gene="crtI"
     CDS             complement(844..1573)
                     /gene="crtI"
                     /note="SCJ12.02c, probable methyltransferase, len: 246 aa.
                     Shares a high level of sequence similarity with
                     Streptomyces griseus TR:P72450 (EMBL: X95596)
                     methyltransferase (242 aa), fasta scores: opt: 984,
                     z-score: 1194.7, E(): 0.65,70% identity in 234 aa overlap.
                     Contains a pfam match to entry PF01209 Ubie_methyltran,
                     ubiE/COQ5 methyltransferase family."
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="putative methyltransferase"
                     /protein_id="CAB53414.1"
                     /db_xref="GI:5777665"
                     /translation="MEIRHHLAAAIDAACETGAFANGTDTRTGLLRAARKLAE
                     AHGAGLHLLLGGGTGASIRALLPAAPRAPITAVDASACMLKPALVKPWPARVRFIH
                     TASEFAIRREDGRPFDAVFAAYLFRNVTDRDAVLRAVRQALTAAHPYSLSSS-
                     GHEAGSGAVTRVVTAILLGLELTHREAGTVAILFAARLTRA---SVRSV-
                     PVAGWGIGIVHTFLARNGSPTTTEERVL"
RBS                  complement(845..850)
                     /gene="crtI"
misc_feature         complement(872..1574)
                     /gene="crtY"
                     /note="Hit/def to entry PF01297 Zn_dep_MoeA,
                     MoeA/MogA family  Score -45.2, E-value
                     1.1e-08"
gene                 complement(1570..2787)
                     /gene="crtY"
CDS                  complement(1570..2787)
                     /gene="crtY"
                     /note="SCJ12.03c, probable lycopene cyclase, len: 405 aa.
                     Highly similar to Streptomyces griseus TR:P72451 (EMBL,
                     X65596) lycopene cyclase (418 aa), fasta scores opt: 1426,
                     z score: 1002.7 E(): 0 56.5% identity in 396 aa overlap."
                     /codon_start=1
                     /transl_table=11
                     /product="putative lycopene cyclase"
                     /protein_id="CAB53415.1"
                     /db_xref="GI:5777666"
                     /translation="MTRSAPLPVLVIVGTGPAALTLAHELHNCTAFAMTLVEPD
                     PLRPAEFIWCYWGAAAIGLEEAVGASWSVLRLHGADGGSVIVDFAPFTYRMVRSADF
                     RSVLDSAPIDAAFLAGTAELVEAVTAGTEVSTLTGRGTTFAARVTLGAGTLEA
                     PAAPFGLLQHEFGWEVETPTDRFDPAVAFLMDTRVPQFAHGLAPGYVLPLAPDRALV
                     TTEFSKAPELFTAVTSAGSRYCKDTTGRGFVVEETEQGVTFMTPARFGPRAPAVVP
                     LTATSGAIRPATGYSFAAVQRHSGATAAALRDGHDRVPAPHGRRARAMDAVLLKALT
                     GRTDGPRFFIDLFRRVPAERLLRFLDGTTSLREEWGTGLRTPVRPMLRTAAEVPTLG
                     RSGLLARTGGNNR"
gene                 complement(4058..4308)
                     /gene="SCJ12.04c"
CDS                  complement(4058..4308)
                     /gene="SCJ12.04c"
                     /note="SCJ12.04c, possible oxidoreductase, len: 416 aa.
                     Similar to many protoporphyrinogen oxidases including:
                     Propionibacterium freudenreichii shermanii SW:PPOX_PROFR
                     (EMBL: D85417) HemY (527 aa), fasta scores opt: 228
                     z score: 249.0 E(): 1.6e-06 26.9% identity in 524 aa
                     overlap and Bacillus subtilis SW:PPOX_BACSU (EMBL: M97208)
                     HemY (470 aa), fasta scores opt: 200, z score: 229.3 E():
                     2.1e-05 25.7% identity in 311 aa overlap. Contains a Pfam
                     match to entry PF01593 Amino_oxidase."
                     /codon_start=1
                     /transl_table=11
                     /product="putative oxidoreductase"
                     /protein_id="CAB53416.1"
                     /db_xref="GI:5777667"
                     /translation="MSIVQHRAAVAVVVAGLAGAAVAIRLEARWPVLLLAAPVL
                     VRMRTLPRRCFLLSGCPLVPNTGYPLVRPEMPLRGLPLRFPTAGVVAHTSNGPVSLAS
                     IRGPRAAPVTPSAALAELSAALAALGSTAVLFVAVTGSSLRLTGAALSVGVL
                     AAIAGTLRPFLSGVFLEDRLRISARFFHLVWRSMVRGSLPLPAEGIGAVPAQLAEGL
                     PDGVLRLGTPVAEITDAGVLLGDGCEVPARVVVVAIDPATAAGLLPDLTVPDERIVT
                     YYHATDSEPCAEPTLMVDSDGATLNTVLSLVADTYAPPGTGLVSISVLTDDPGSAQ
                     AVLELAEQVIGIDRSWEQVAAPTVCRALPAMCDFWELSPTTRIGPCHYVGDHRNAIG
                     SVQGAIASGIRAARRVRADKIPGR"
misc_feature         complement(4082..4247)
                     /gene="SCJ12.04c"
                     /note="Hit/def to entry PF01593 Amino_oxidase, score
                     -125.60, E-value 0.0012"
RBS                  complement(4414..4419)
gene                 complement(4542..5541)
                     /gene="SCJ12.05c"
CDS                  complement(4542..5541)
                     /gene="SCJ12.05c"
                     /note="SCJ12.05c, possible DNA binding regulator, len: 333
                     aa. Some similarity to Myxococcus xanthus TR:Q50900 (EMBL,
```

```
                     Z...95) hypothetical protein, fasta scores: opt: ...2
                     z score: 1...4 E(): 1...-05 26.1% identity in 280 aa overlap
                     and MyxA...ein, xanthus ... (EMBL: ... 1.2e-05) fasta scores:
                     opt: 1...2 z score: 1...4 E(): 1...-05 26.1% identity in 280
                     aa overlap. Contains a ...
                     /codon_start=1
                     /transl_table=11
                     /product="putative DNA ... regulator"
                     /protein_id="CAB53417.1"
                     /db_xref="GI:5777668"
                     /translation="MSGPVSLTLLGPVEVRVDGRPVPLGGPKQRTVLAVLALSA
                     GRAVSTDRLVDALWGEEPPATARNALQVYVSQLRKALEPDRARGAPSRVLVTEPPGY
                     MLRVEPEEVDAVRFERLVAEGRRALAAGDPAAAAGVLREALALWRGPPLADVAYEPF
                     AQAEIARLEELRLVAVEDRIDADLALGRHAEVVGELEALVAEHPLRERLRGQLMLAL
                     YRSGRQADALEAYQRARRTLVDELGIEPGPELRELERRILAQDPSLDAPAPAPAPAP
                     AAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP
                     AP
```

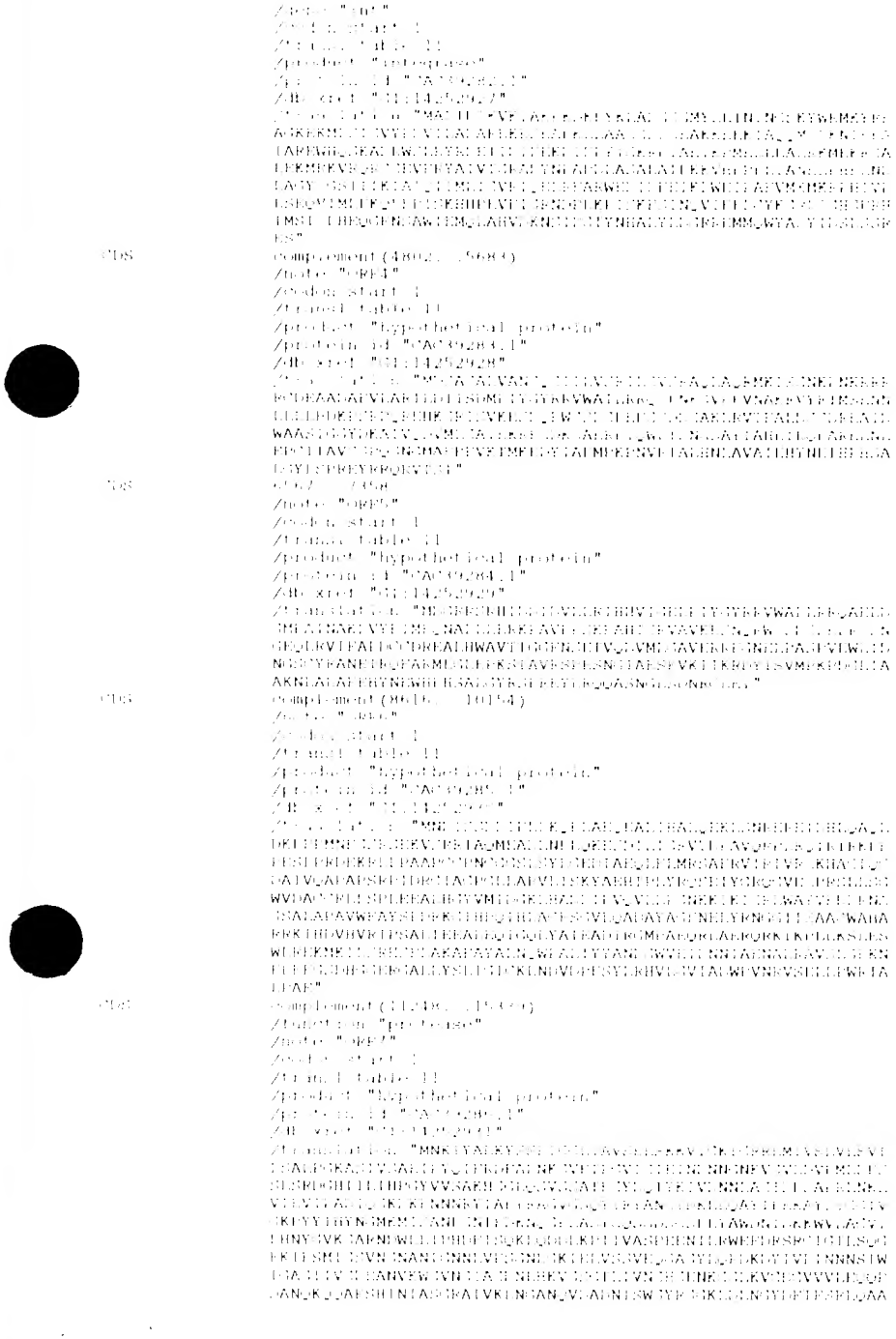
```
            Web site: http://www.seq.wi.mit.edu
            Contact: sequence-submissions@genome.wi.mit.edu
                     Project Information
            Center project name: 10693
            Center clone name: 406_N_10
                     Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Cloning: bye termination kit bye; 100% of reads
            Assembly programs: Phrap version 0.960731
            Consensus quality: 16,202 bases at least q40
            Consensus quality: 17064 bases at least q30
            Consensus quality: 17905 bases at least q20
            Unsolved gaps: 18,200; advanse tp
            Unsolved size: 17628; sum of contigs
            Quality coverage: 3.5 in q20 bases; advanse tp
            Quality coverage: 3.6 in q20 bases; sum of contigs
                     .
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*         1    1129: contig of 1129 bp in length
*      1130 1229: gap of       100 bp
*      1230    2475: contig of 1146 bp in length
*      2476 2575: gap of       100 bp
*      2576      4103: contig of 1528 bp in length
*      4104 4203: gap of       100 bp
*      4204    6767: contig of 2564 bp in length
*      6768 6867: gap of       100 bp
*      6868      8323: contig of 1456 bp in length
*      8324 8423: gap of       100 bp
*      8424    10157: contig of 1734 bp in length
*     10158 10257: gap of       100 bp
*     10258    12751: contig of 2494 bp in length
*     12752 12851: gap of       100 bp
*     12852    14865: contig of 2014 bp in length
*     14866 14965: gap of       100 bp
*     14966    16709: contig of 1744 bp in length
*     16710 16809: gap of       100 bp
*     16810    20004: contig of 3195 bp in length
*     20005 20104: gap of       100 bp
*     20105    24067: contig of 4163 bp in length
*     24068 24167: gap of       100 bp
*     24168    27387: contig of 3220 bp in length
*     27388 27487: gap of       100 bp
*     27488    29787: contig of 2300 bp in length
*     29788 29887: gap of       100 bp
*     29888    33967: contig of 4080 bp in length
*     33968 34067: gap of       100 bp
*     34068    38067: contig of 4000 bp in length
*     38068 38167: gap of       100 bp
*     38168    42441: contig of 4274 bp in length
*     42442 42541: gap of       100 bp
*     42542    46821: contig of 4280 bp in length
*     46822 46921: gap of       100 bp
*     46922    53521: contig of 6600 bp in length
*     53522 53621: gap of       100 bp
*     53622    60245: contig of 6624 bp in length
*     60246 60345: gap of       100 bp
*     60346    63057: contig of 2712 bp in length
*     63058 63157: gap of       100 bp
*     63158    71957: contig of 8800 bp in length
*     71958 72057: gap of       100 bp
*     72058    80857: contig of 8800 bp in length
*     80858 80957: gap of       100 bp
*     80958    89757: contig of 8800 bp in length
*     89758 89857: gap of       100 bp
*     89858    100641: contig of 10784 bp in length
*    100642 100741: gap of       100 bp
*    100742    112285: contig of 11544 bp in length
*    112286 112385: gap of       100 bp
*    112386    126665: contig of 14280 bp in length
*    126666 126765: gap of       100 bp
*    126766    141518: contig of 14753 bp in length
*    141519 141618: gap of       100 bp
*    141619    157298: contig of 15680 bp in length
*    157299 157398: gap of       100 bp
*    157399    178988: contig of 21588 bp in length.
FEATURES             Location/Qualifiers
     source          1..178988
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-406N10"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1129
                     /note="assembly_fragment"
     misc_feature    1130..1229
                     /note="assembly_fragment"
     misc_feature    1230..2475
                     /note="assembly_fragment"
     misc_feature    2476..2575
                     /note="assembly_fragment"
     misc_feature    2576..4103
                     /note="assembly_fragment"
     misc_feature    4104..4203
                     /note="assembly_fragment"
                     clone_end='
                     vector_side=right"
     misc_feature    8424..10157
                     /note="assembly_fragment"
     misc_feature    10258..12751
                     /note="assembly_fragment"
     misc_feature    12852..14865
                     /note="assembly_fragment"
     misc_feature    14966..16709
                     /note="assembly_fragment"
     misc_feature    17610..20004
                     /note="assembly_fragment"
     misc_feature    20105..24067
                     /note="assembly_fragment"
     misc_feature    25009..27387
                     /note="assembly_fragment"
     misc_feature    27899..29787
                     /note="assembly_fragment"
     misc_feature    31784..33967
                     /note="assembly_fragment"
     misc_feature    35963..38067
                     /note="assembly_fragment"
     misc_feature    40170..42441
                     /note="assembly_fragment"
     misc_feature    44644..46821
                     /note="assembly_fragment"
     misc_feature    49824..53521
                     /note="assembly_fragment"
     misc_feature    56030..60245
                     /note="assembly_fragment"
     misc_feature    62754..63057
                     /note="assembly_fragment"
     misc_feature    71866..80723
                     /note="assembly_fragment"
     misc_feature    80824..89757
                     /note="assembly_fragment"
     misc_feature    89758..100641
                     /note="assembly_fragment"
     misc_feature    100842..112285
                     /note="assembly_fragment"
     misc_feature    112286..126665
                     /note="assembly_fragment"
     misc_feature    126766..141518
                     /note="assembly_fragment"
     misc_feature    141619..157298
                     /note="assembly_fragment"
     misc_feature    157399..178988
                     /note="assembly_fragment"
```

```
                    clone_end:SP6
                    vector_side:left"
BASE COUNT    48479 a   43855 c   39079 g   49372 t    2295 others
ORIGIN


  Query Match             53.8%;  Score 22.8,  DB 2,  Length 178986,
  Best Local Similarity   71.4%;  Pred. No. 5%;
  Matches   20;  Conservative    0,  Mismatches   12,  Indels    0,  Gaps    0;

Qy       2 attcatcagcgacactcaggcctcacccgattgataactcaa 43
           |:|||| |||| |:|||||| ||| |:|  |||||||||||
Db   93897 GTTCTTCCCCCACCCTGCTGCCCCCACTGCTGTCCAGCTGCA 93938


RESULT   6
AP003100
ID   AP003100   192754 bp    DNA             HTG       18-JAN-2001
DEFINITION  Homo sapiens chromosome 11 clone RP11-27G22 map 11q, WORKING DRAFT
            SEQUENCE, 29 unordered pieces.
ACCESSION   AP003100
VERSION     AP003100.1  GI:12275507
KEYWORDS    HTG; HTGS_BASEL; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-27G22.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 192754)
  AUTHORS   Hattori,M., Ishii,K., Iwayoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens DNA of 11q
  JOURNAL   Published Only in DataBase (2001) In press
REFERENCE   2  (bases 1 to 192754)
  AUTHORS   Hattori,M., Ishii,K., Iwayoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-2001) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:[illegible], Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT                 Genome Center
        Center: RIKEN Genomic Sciences Center(GSC)
        Center code: RIKEN
        Web site: http://[illegible]
        Contact: hattori@gsc.riken.go.jp
        ----------  Project Information
        Center project name: HumDraft11
        Center clone name: RP11-27G22
        ----------  Summary Statistics
        Sequencing vector: PCR products, 100% of reads
        Chemistry: Dye-terminator ET-amersham; 100% of reads
        Assembly program: Phrap; version 0.990329
        Consensus quality: 185694 bases at least Q40
        Consensus quality: 188490 bases at least Q30
        Consensus quality: 189463 bases at least Q20
        Insert size: 189953; sum-of-contigs
        Quality coverage: 9.45x in Q20 bases, sum = 1 contigs
        ----------
NOTE: This is a 'working draft' sequence. It currently consists of
29 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved.
          1      17915 contig of   17915 bp in length
      18016      32894 contig of   14879 bp in length
      32995      45749 contig of   12755 bp in length
      45850      59508 contig of   13659 bp in length
      59609      72172 contig of   12564 bp in length
      72273      83540 contig of   11268 bp in length
      83641      93073 contig of    9433 bp in length
      93174     102238 contig of    9065 bp in length
     102339     113626 contig of   11288 bp in length
     113727     118548 contig of    4822 bp in length
     118649     124803 contig of    6155 bp in length
     124904     130263 contig of    5360 bp in length
     130364     136333 contig of    5970 bp in length
     136434     142507 contig of    6074 bp in length
     142604     143832 contig of    1229 bp in length
     143933     145125 gap of       1293 bp
     145226     153229 contig of    8004 bp in length
     153330     158063 contig of    4734 bp in length
     158163     162812 contig of    4650 bp in length
     162913     166611 contig of    3699 bp in length
     166712     167617 contig of     906 bp in length
     167718     173153 contig of    5436 bp in length
     173254     173432 gap of        179 bp in length
     173533     178433 contig of    4901 bp in length
     178534     180882 contig of    2349 bp in length
     180983     182506 contig of    1524 bp in length
     182607     186092 contig of    3486 bp in length
     186193     187637 contig of    1445 bp in length
     187738     189217 contig of    1480 bp in length
     189318     190634 contig of    1317 bp in length
     190735     192258 contig of    1524 bp in length
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*          1      17915 contig of   17915 bp in length
*      17916      18015 gap of        100 bp
*      18016      32894 contig of   14879 bp in length
*      32895      32994 gap of        100 bp
*      32995      45749 contig of   12755 bp in length
*      45750      45849 gap of        100 bp
*      45850      59508 contig of   13659 bp in length
*      59509      59608 gap of        100 bp
*      59609      72172 contig of   12564 bp in length
*      72173      72272 gap of        100 bp
*      72273      83540 contig of   11268 bp in length
*      83541      83640 gap of        100 bp
*      83641      93073 contig of    9433 bp in length
*      93074      93173 gap of        100 bp
*      93174     102238 contig of    9065 bp in length
*     102239     102338 gap of        100 bp
*     102339     113626 contig of   11288 bp in length
*     113627     113726 gap of        100 bp
*     113727     118548 contig of    4822 bp in length
*     118549     118648 gap of        100 bp
*     118649     124803 contig of    6155 bp in length
*     124804     124903 gap of        100 bp
*     124904     130263 contig of    5360 bp in length
*     130264     130363 gap of        100 bp
*     130364     136333 contig of    5970 bp in length
*     136334     136433 gap of        100 bp
*     136434     142507 contig of    6074 bp in length
*     142508     142603 gap of         96 bp
*     142604     143832 contig of    1229 bp in length
*     143833     145125 gap of       1293 bp
*     145226     153229 contig of    8004 bp in length
*     153330     158063 contig of    4734 bp in length
*     158064     158162 gap of         99 bp
*     158163     162812 contig of    4650 bp in length
*     162813     162912 gap of         100 bp
*     162913     166614 contig of    3702 bp in length
*     166615     166713 gap of         99 bp
*     166714     167617 contig of     904 bp in length
*     167618     167717 gap of         100 bp
*     167718     173153 contig of    5436 bp in length
*     173154     173253 gap of         100 bp
*     173254     173432 contig of     179 bp in length
```

```
     *    [illegible] [illegible]: gap of       [illegible] bp
     *    [illegible]   [illegible]: contig of  [illegible] bp in length
     *    [illegible] [illegible]: gap of       [illegible] bp
     *    [illegible]   [illegible]: contig of  [illegible] bp in length
     *    [illegible] [illegible]: gap of       [illegible] bp
     *    [illegible]   [illegible]: contig of  [illegible] bp in length
     *    [illegible] [illegible]: gap of       [illegible] bp
     *    [illegible]   [illegible]: contig of  [illegible] bp in length
     *    [illegible] [illegible]: gap of       [illegible] bp
     *    [illegible]   [illegible]: contig of  [illegible] bp in length
     *    [illegible] [illegible]: gap of       [illegible] bp
     *    [illegible]   [illegible]: contig of  [illegible] bp in length
     *    [illegible] [illegible]: gap of       [illegible] bp
     *    [illegible]   [illegible]: contig of  [illegible] bp in length
     *    [illegible] [illegible]: gap of       [illegible] bp
     *    [illegible]   [illegible]: contig of  [illegible] bp in length.
FEATURES             Location/Qualifiers
     source          1..192754
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-27G22"
     misc_feature    1..17415
                     /note="assembly_fragment"
     misc_feature    18016..32894
                     /note="assembly_fragment"
     misc_feature    32895..45744
                     /note="assembly_fragment"
     misc_feature    45855..55508
                     /note="assembly_fragment"
     misc_feature    56005..72172
                     /note="assembly_fragment clone end:SP6 vector_side:left"
     misc_feature    72273..88040
                     /note="assembly_fragment"
     misc_feature    88041..93073
                     /note="assembly_fragment"
     misc_feature    93174..102220
                     /note="assembly_fragment"
     misc_feature    102331..113626
                     /note="assembly_fragment"
     misc_feature    113727..118948
                     /note="assembly_fragment"
     misc_feature    118949..124803
                     /note="assembly_fragment"
     misc_feature    124904..130262
                     /note="assembly_fragment"
     misc_feature    130363..136238
                     /note="assembly_fragment"
     misc_feature    136341..142503
                     /note="assembly_fragment"
     misc_feature    142604..148328
                     /note="assembly_fragment"
     misc_feature    148429..153128
                     /note="assembly_fragment"
     misc_feature    153229..158062
                     /note="assembly_fragment"
     misc_feature    158163..166812
                     /note="assembly_fragment"
     misc_feature    166913..166914
                     /note="assembly_fragment"
     misc_feature    166915..167967
                     /note="assembly_fragment clone end:T7 vector_side:left"
     misc_feature    168068..173350
                     /note="assembly_fragment"
     misc_feature    173451..178432
                     /note="assembly_fragment"
     misc_feature    178433..180082
                     /note="assembly_fragment"
     misc_feature    180083..182506
                     /note="assembly_fragment"
     misc_feature    182607..186442
                     /note="assembly_fragment"

Query Match          54.0%;  Score 22.8;  DB 2;  Length 192754;
```

```
  Best Local Similarity   71.4%;  Pred. No. 57;
  Matches   40;  Conservative    3;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       2 attcgtcaaccacacctactgacctcaccagttgataggctaca 43
           |||| ||  | | | | |  ||  |||||| |  |||||||
Db   45173 GTTCTTCTCCCACCCTGTTCCCCTCACTGGTGTCCAGCTGCA 45214


RESULT   7
AC022387
LOCUS       AC022387   190016 bp    DNA             HTG       22-JUL-2000
DEFINITION  Homo sapiens clone RP11 27G22, WORKING DRAFT SEQUENCE, 15 unordered
            pieces.
ACCESSION   AC022387
VERSION     AC022387.4   GI:6580478
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 190016)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-27G22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 190016)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,B., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bedu,F.,
            Boqunlavkiy,L., Bouhqalter,B., Brown,A., Burkett,G., Castle,A.,
            Choopel,Y., Coianqelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeAreliano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzBuh,W., Forrest,C., Gane,D., Galaqan,J.,
            Gardyna,S., Grant,G., Jaqos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,C., Jones,C., Kann,L., Kuratas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Lin,G., Locke,K.,
            Macdonald,P., Marques,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Neiman,C.H., O'Donnell,K., O'Donnell,P., Oliver,I.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Savery,P., Spencer,B., Stabac-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 22, 2000 this sequence version replaced gi:7249094.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            [illegible]
            --- -------- Genome Center
                Center: Whitehead Institute/MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence-submissions@genome.wi.mit.edu
            - --- ------ Project Information
                Center project name: LA691
                Center clone name: 27_G_22
            -          Summary Statistics
                Sequencing vector: 2 M13: M77815: 97% of reads
                Sequencing vector: Plasmid: n/a: 3 0.198% of reads
                3.5:1769:2247:1 Chemistry: Dye Terminator Big Dye: 100% of
            reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 190834 bases at least Q40
                Consensus quality: 194422 bases at least Q30
                Consensus quality: 195799 bases at least Q20
                Insert size: 140000; agarose-tp
                Insert size: 157615; sum-of-contigs
                Quality coverage: 5.4 in Q20 bases; agarose-tp
                Quality Cov.
            * NOTE: This is a "working draft" sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
```

```
            * arbitrary  Gaps between the contigs are represented as
            * runs of N  (i.e. the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1127: contig of 1127 bp in length
            *     1128  1227: gap of      100 bp
            *     1228     3640: contig of 2413 bp in length
            *     3641  3740: gap of      100 bp
            *     3741     7692: contig of 3952 bp in length
            *     7693  7792: gap of      100 bp
            *     7793     12634: contig of 4842 bp in length
            *    12635  12734: gap of      100 bp
            *    12735     16941: contig of 4207 bp in length
            *    16942  17041: gap of      100 bp
            *    17042     21720: contig of 4679 bp in length
            *    21721  21820: gap of      100 bp
            *    21821     27287: contig of 5467 bp in length
            *    27288  27387: gap of      100 bp
            *    27388     31896: contig of 4509 bp in length
            *    31897  31996: gap of      100 bp
            *    31997     36686: contig of 4690 bp in length
            *    36687  36786: gap of      100 bp
            *    36787     49741: contig of 12955 bp in length
            *    49742  49841: gap of      100 bp
            *    49842     72807: contig of 22966 bp in length
            *    72808  72907: gap of      100 bp
            *    72908     94912: contig of 22005 bp in length
            *    94913  95012: gap of      100 bp
            *    95013    124413: contig of 29401 bp in length
            *   124414  124513: gap of      100 bp
            *   124514    162373: contig of 37860 bp in length
            *   162374  162473: gap of      100 bp
            *   162474    199016: contig of 36543 bp in length.
FEATURES             Location/Qualifiers
     source          1..199016
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-27G22"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1127
                     /note="assembly_fragment"
     misc_feature    1228..3640
                     /note="assembly_fragment"
     misc_feature    3741..7692
                     /note="assembly_fragment"
     misc_feature    7793..12634
                     /note="assembly_fragment"
     misc_feature    12735..16941
                     /note="assembly_fragment"
     misc_feature    17042..21720
                     /note="assembly_fragment"
     misc_feature    21821..27287
                     /note="assembly_fragment"
     misc_feature    27388..31896
                     /note="assembly_fragment"
     misc_feature    31997..36686
                     /note="assembly_fragment"
     misc_feature    36787..49741
                     /note="assembly_fragment"
     misc_feature    49842..72807
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    72908..94912
                     /note="assembly_fragment"
     misc_feature    95013..124413
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    124514..162373
                     /note="assembly_fragment"
     misc_feature    162474..199016
                     /note="assembly_fragment"
BASE COUNT     55479 a  44052 c  44 g  55008 t     104 others
ORIGIN


  Query Match             54.0%   Score 4220   DB      Length 199016;
  Best Local Similarity   97.4%   Pred. No.
  Matches    60;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 gttcatcagccacctatcagtcgttcaatcctcttatc  43
           |||| || |||| || |||  ||| ||| |||  |||
Db   83957 GTTCTTCGCCTACCTTCTCGCTCTTCATCGTCTTTC 83914


RESULT   8
EC0278144/c
LOCUS       EC0278144    4771 bp    DNA             BCT       22-MAY-2
DEFINITION  Escherichia coli partial genomic DNA,  pathogenicity island,
            shiga toxin producing strain 4797/97.
ACCESSION   AJ278144
VERSION     AJ278144.1  GI:4252
KEYWORDS    in' gene, intergenic, SPI    SE,    S    ,    -PI4; orf1; orf2;
            ORF17; ORF 8; ORF19; ORF4   ORF   ORF   ORF7, ORF6; ORF9; putative
            sodium galactoside symport
SOURCE      Escherichia co...
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gamma  ...  ...  Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases  to 4771)
  AUTHORS   Zhang,W.L., Bielaszewska,M.  ...  E.,  ...  and Karch,H.
  TITLE     Characterisation of a novel pathogenicity island present in
            EHEC/enteropathogenic Shiga toxin-producing Escherichia coli
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4771)
  AUTHORS   Zhang,W.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-  ) Zhang W.  .  ., Max Von Pettenkofer-Inst
            Mikrobiologie, Universitaet Muenchen, Pettenkofer-Str. 2,
            D-80336, GERMANY
FEATURES             Location/Qualifiers
     source          1..4771
                     /organism="Escher  ...  l"
                     /strain="4797/97"
                     /serotype="  ...  "
                     /db_xref="taxon:562"
                     /country="Germany"
                     /note="STEC from an asymptomatic carrier"
     gene            complement(2..2...  2604
                     /gene="yi1"
     CDS             complement(2..219...  2604
                     /gene="yi1"
                     /codon_start=1
                     /transl_table=11
                     /product="putative  ...  ...  ...  ...protease/importer"
                     /protein_id="CAC...281.1"
                     /db_xref="GI:14...502"
                     /translation="MK...  ...  ...  ...  ...  ...  ...  ...
                     ...  ...  ...  
```



/gene="tnf"
/codon_start=1
/transl_table=11
/product="interleukin-8"
/protein_id="AAC39283.1"
/db_xref="GI:1425929"
/translation="MACTTQMAVFTTKRSRVAAWCAALFLVSMLTMDEGVFEAAIESIINVFHQYSVKEGHFDVLKNEQLLVKIGKIFDAIIQDTIGGMAQEMVSEIKTVAVLSATLNFHMREGNLEQIKRNYQSMAFRDRDADMESQKAFKMSREKLHEECPCVETSVTNTIHDNFQKCLISNISEVVHRGCGCNF"
CDS complement(4802..5683)
/note="ORF4"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC39284.1"
/db_xref="GI:1425929"
/translation="MVLAVADVAAVLLVSVSLVVLAVLAMEGLSRVIHEGQDFALLGIKVNDIQNVVFASFGLVLVAAVYQTFFTSKISSESIRREIYRQIQHFTSEMRTQGTTISANLKDENSRVSEQIVQTFGQNITTLHTFIASFIGGFVIGFIKGWLLTLVMLSSIPVLAIAGGVMAIMVSKMASRGQAAYAKAGAVAEEVLGAIRTVIAFGGQKKELERYNKNLEEAKRIGIKKAITANISMGAAFLLIYASYALAFWYGTSLVISKEYTIGQVLTVFFSVLIGAFSVGQASPSIEAFANARGAAYEIFKIIDNKPSIDSYSKSGHKPDNIKGNLEFRNVHFSYPSRKEVQILKGLNLKVQSGQTVALVGNSGCGKSTTVQLMQRLYDPTEGMVSIDGQDIRTINVRFLREIIGVVSQEPVLFATTIAENIRYGREDVTMDEIEKAVKEANAYDFIMKLPHQFDTLVGERGAQLSGGQKQRIAIARALVRNPKILLLDEATSALDTESEKVVQEALDKARKGRTTIVIAHRLSTIQNADLIVVIQNGKVKEHGTHQQLLAQKGIYFSMVSVQAGTKR" /protein_id="AAC39284.1" /db_xref="GI:1425929" /translation="MEVLAVADVAAVLLVSVSLVVLAVLAMEGLSRVIHEGQDFALLGIKVNDIQNVVFASFGLVLVAAVYQTFFTSKISSESIRREIYRQIQHFTSEMRTQGTTISANLKDENSRVSEQIVQTFGQNITTLHTFIASFIGGFVIGFIKGWLLTLVMLSSIPVLAIAGGVMAIMVSKMASRGQAAYAKAGAVAEEVLGAIRTVIAFGGQKKELERYNKNLEEAKRIGIKKAITANISMGAAFLLIYASYALAFWYGTSLVISKEYTIGQVLTVFFSVLIGAFSVGQASPSIEAFANARGAAYEIFKIIDNKPSIDSYSKSGHKPDNIKGNLEFRNVHFSYPSRKEVQILKGLNLKVQSGQTVALVGNSGCGKSTTVQLMQRLYDPTEGMVSIDGQDIRTINVRFLREIIGVVSQEPVLFATTIAENIRYGREDVTMDEIEKAVKEANAYDFIMKLPHQFDTLVGERGAQLSGGQKQRIAIARALVRNPKILLLDEATSALDTESEKVVQEALDKARKGRTTIVIAHRLSTIQNADLIVVIQNGKVKEHGTHQQLLAQKGIYFSMVSVQAGTKR"
CDS 6567..7358
/note="ORF5"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC39284.1"
/db_xref="GI:1425929"
/translation="MRGFRFHLHSSVVLLKSHRVIHEEIYYGYRVWAIEGLAELD
SMLAIGALVVFLMGALNLKEAVFVGTAFLAHQVAVELGWLVFVGREIQVAMTAVG
GMGYLRVTEAFDQVDREALHWAVTIGGFNLTFVGLVMLSAVEKFGNGLVAGTFVLW
NGSFYFANETLQLPAKMLSLFESSIAVESVLSNLIAESFVKLTKRVGLSVMLKPG
AKNLALAFFRYNEWHLESALSYFLPFFTLGQASNGLLSNMLVTL"
CDS complement(8661..10154)
/note="ORF6"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC39285.1"
/db_xref="GI:1425930"
/translation="MRLSLLFLLAELLAVLALHAAERLNREEEEEEDRGAAL
DRTLFMRLSLSLEEVLPSIAQMEALLNELQELPLPLSVVLAVQFVLSVFTKEEEE
EESLPRDERELLPAADVVDPNVQSSIYYCGEIAFQLEFMRSAFNKVFVRKHATEV
GATVQAFAPSRPLDMCIAQMPCLLAFVLLPKYAEHTPLYIPQPFLYDRQVVEC
WVDAVCPLLSPLEEALRETVMPLGRLRAGLTLEVLVLGSRREKEIGSWAVVEFLN
GSALAFAVWPAYSLDFKSIHDWLTHLAVFSVVLAIAYAYAGNELYKNGTLLAAVWAHA
RKRTHDVHVKLTNSALTEEAFDKFGQQLYATRAQTRGMEARQRLAREQRKLKELGKSLES
WLREKMEVSP

```
  Matches   29;  Conservative    0;  Mismatches    11;  Indels    0;  Gaps    0;

Qy          3 ttcatcaaacaacgctattgacctaaccattataagctgc 42
              ||||||| | | |||||||   | ||||  | | | |||
Db       7049 TCTGTCAATCACCATTATGGAGACGCCGGAAACTCGTTGC 7006


RESULT  9
RP341E19/C
LOCUS       RP341E18    92558 bp    DNA             PRI       04-APR-2001
DEFINITION  Human DNA sequence from clone RP3-341E18 on chromosome 6p11.2-12.3
            Contains the KIAA0936 gene, for a MAK related kinase and a novel
            alternatively spliced gene. Contains a putative CpG island, ESTs
            and GSSs, complete sequence.
ACCESSION   AL031178
VERSION     AL031178.1  GI:3650040
KEYWORDS    HTG; CpG island; KIAA0936; MAK-related kinase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 92558)
  AUTHORS   Tubby,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA  UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 6, 1998 this sequence version replaced gi:3421037.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted, all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality
            40); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats, all regions were covered by at least
            one plasmid subclone or more than one M13 subclone, and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep.shtml.
            This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/Chr6/
            RP3-341E18 is from the library RPCI-3 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/
            VECTOR: pCYPAC2
            This sequence is the entire insert of clone RP3-341E18.
FEATURES             Location/Qualifiers
     source          1..92558
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /chromosome="6"
                     /map="p11.2-12.3"
                     /clone="RP3-341E18"
                     /clone_lib="RPCI-3"
     repeat_region   5648..5673
                     /note="13 copies 2 mer bp 96% conserved"
     repeat_region   5682..5767
                     /note="43 copies 2 mer at 62% conserved"
     misc_feature    complement(5768..6046)
                     /note="match: GSS: Em:AQ126650"
     misc_feature    complement(5769..5806)
                     /note="match: STS: Em:HS107Y8"
     misc_feature    complement(5769..5808)
                     /note="match: STS: Em:HS015X5"
     misc_feature    complement(5769..5841)
                     /note="match: STS: Em:HSA[illegible] Em:HSA[illegible] Em:HSA[illegible]
                     Em:HSA[illegible]YA5 Em:HUMPEP[illegible]
                     match: GSS: Em:AQ[illegible] Em:AQ[illegible] Em:AQ[illegible]
                     Em:AQ[illegible]"
     misc_feature    5770..5803
                     /note="match: STS: Em:HS[illegible]"
     misc_feature    5770..5814
                     /note="match: STS: Em:HUMPEP[illegible]
                     match: GSS: Em:AQ[illegible] Em:AQ[illegible]"
     misc_feature    5772..5805
                     /note="match: STS: Em:HS[illegible]"
     misc_feature    5772..5814
                     /note="match: GSS: Em:AQ[illegible]"
     misc_feature    5772..5903
                     /note="match: GSS: Em:A[illegible]"
     misc_feature    5773..5905
                     /note="match: GSS: Em:A[illegible]"
     misc_feature    5774..5812
                     /note="match: GSS: Em:A[illegible]"
     misc_feature    5826..5852
                     /note="match: GSS: Em:A[illegible]"
     misc_feature    complement(5826..5844)
                     /note="match: GSS: Em:A[illegible]"
     misc_feature    complement(5826..5847)
                     /note="match: GSS: Em:A[illegible]"
     misc_feature    complement(5826..5848)
                     /note="match: GSS: Em:A[illegible]"
     misc_feature    5871..5894
                     /note="match: GSS: Em:A[illegible]"
     mRNA            complement(join(5893..6046,7960..8064,14754..14855,17960..19267))
                     /gene="dJ341E18.2"
                     /note="match: cDNAs: Em:[illegible] Em:X60[illegible] Em:M3587,"
                     /evidence=not_experimental
                     /product="[illegible]kinase"
     gene            complement(5893..18464)
                     /gene="dJ341E18.2"
     CDS             complement(join(5893..5973,6041..6064,14754..14855,
                     17960..18464))
                     /gene="dJ341E18.2"
                     /note="contained in AL[illegible].12. AL[illegible]
                     [illegible]
                     SW:P43294 1:1409"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="dJ341E18.1 (KIAA[illegible] (MAK related kinase)"
                     /protein_id="CAA[illegible]"
                     /db_xref="GI:3684[illegible]"
                     /db_xref="SPTREMBL:O75[illegible]"
                     /translation="MNLSVLQLFLSNSLAIKPVKIKLSWEDIMNLSLVKLCNDSVVELA
                     KNIMYQIIQSIAFIHKHG"
     repeat_region   9150..9283
                     /note="LINE1 repeats [illegible] 6.0% 1 consensus"
     repeat_region   9234..9[illegible]
                     /note="2 copies [illegible] 96% conserved"
     misc_feature    complement(9[illegible]..9[illegible]
                     /note="m34 EST"
                     /note="match: GSS: Em:[illegible]"
     misc_feature    9[illegible]..9[illegible]
                     /note="match: GSS: Em:[illegible]"
     misc_feature    complement(10[illegible]..10[illegible])
```

```
                     /note= "L1541F18.1"
                     /db_xref= "match: SIS: Em:G20023"
     repeat_region   184..[illegible]
                     /note= "MER5A repeat match: 1..[illegible] consensus"
     misc_feature    2566..2615
                     /note= "match: GSS: Em:AQ372247"
     misc_feature    complement(2716..2778)
                     /note= "match: GSS: Em:AQ388043"
     misc_feature    2734..2796
                     /note= "match: GSS: Em:AQ[illegible]"
     misc_feature    2748..2796
                     /note= "match: SIS: Em:BG[illegible]"
     misc_feature    2816..2862
                     /note= "match: GSS: Em:AQ480818"
     misc_feature    complement(2849..2862)
                     /note= "match: GSS: Em:AQ[illegible]"
                     match: SIS: Em:HSAL08WE1"
     misc_feature    complement(2850..2884)
                     /note= "match: GSS: Em:AQ317636"
     misc_feature    complement(2850..2896)
                     /note= "match: GSS: Em:AQ586321"
     misc_feature    complement(2850..2881)
                     /note= "match: GSS: Em:AQ446[illegible]"
     misc_feature    2850..2896
                     /note= "match: GSS: Em:AQ526760"
     misc_feature    2881..2904
                     /note= "match: SIS: Em:[illegible]"
     misc_feature    complement(2925..2979)
                     /note= "match: SIS: Em:G21741"
     misc_feature    2955..2983
                     /note= "match: SIS: Em:H11427"
     repeat_region   2973..3012
                     /note= "L1HA repeat match 5761..6113 of consensus"
     repeat_region   3202..3776
                     /note= "L1 Hs repeat from ap [illegible] consensus"
     misc_feature    complement(3787..3804)
                     /note= "match: GSS: Em:AQ461611"
     misc_feature    3811..3865
                     /note= "cpG island"
                     /gene= "CpG: [illegible]"
     misc_feature    complement(4036..4084)
                     /note= "match: GSS: Em:AQ311072"
     misc_feature    complement(4044..4084)
                     /note= "match: GSS: Em:AQ150229"
     misc_feature    4208..4270
                     /note= "cpG island"
                     /gene= "CpG: [illegible]"
     mRNA            join(4234..4302,4788..4867,5033..5111,5046[illegible],5332..5337[illegible],
                     [illegible],6[illegible],6[illegible],6[illegible],6[illegible],6[illegible],
                     6[
```

```
*        1     1142: contig of 1142 bp in length
*     1143  1242: gap of      100 bp
*     1243     2438: contig of 1196 bp in length
*     2439  2538: gap of      100 bp
*     2539     3714: contig of 1176 bp in length
*     3715  3814: gap of      100 bp
*     3815     5559: contig of 1745 bp in length
*     5560  5659: gap of      100 bp
*     5660     7245: contig of 1586 bp in length
*     7246  7345: gap of      100 bp
*     7346     9329: contig of 1984 bp in length
*     9330  9429: gap of      100 bp
*     9430    11234: contig of 1805 bp in length
*    11235 11334: gap of      100 bp
*    11335    12701: contig of 1367 bp in length
*    12702 12801: gap of      100 bp
*    12802    14171: contig of 1370 bp in length
*    14172 14271: gap of      100 bp
*    14272    15608: contig of 1337 bp in length
*    15609 15708: gap of      100 bp
*    15709    17232: contig of 1524 bp in length
*    17233 17332: gap of      100 bp
*    17333    18977: contig of 1645 bp in length
*    18978 19077: gap of      100 bp
*    19078    21304: contig of 2227 bp in length
*    21305 21404: gap of      100 bp
*    21405    25252: contig of 3848 bp in length
*    25253 25352: gap of      100 bp
*    25353    28561: contig of 3209 bp in length
*    28562 28661: gap of      100 bp
*    28662    31363: contig of 2702 bp in length
*    31364 31463: gap of      100 bp
*    31464    35304: contig of 3841 bp in length
*    35305 35404: gap of      100 bp
*    35405    39966: contig of 4562 bp in length
*    39967 40066: gap of      100 bp
*    40067    42541: contig of 2475 bp in length
*    42542 42641: gap of      100 bp
*    42642    46458: contig of 3817 bp in length
*    46459 46558: gap of      100 bp
*    46559    51258: contig of 4700 bp in length
*    51259 51358: gap of      100 bp
*    51359    54202: contig of 2844 bp in length
*    54203 54302: gap of      100 bp
*    54303    59459: contig of 5157 bp in length
*    59460 59559: gap of      100 bp
*    59560    63206: contig of 3647 bp in length
*    63207 63306: gap of      100 bp
*    63307    67217: contig of 3911 bp in length
*    67218 67317: gap of      100 bp
*    67318    72748: contig of 5431 bp in length
*    72749 72948: gap of      100 bp
*    72849    78863: contig of 6015 bp in length
*    78864 78963: gap of      100 bp
*    78964    84076: contig of 5113 bp in length
*    84077 84176: gap of      100 bp
*    84177    92283: contig of 8107 bp in length
*    92284 92383: gap of      100 bp
*    92384   101924: contig of 9441 bp in length
*   101825 101924: gap of      100 bp
*   101925   111865: contig of 9941 bp in length
*   111866 111965: gap of      100 bp
*   111966   122994: contig of 11029 bp in length
*   122995 123094: gap of      100 bp
*   123095   139594: contig of 16500 bp in length.
FEATURES             Location/Qualifiers
     source          1. .139594
                     /organism="Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome="6"
                     /map "6"
                     /clone="RP11-163013"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1142
                     /note "assembly_fragment"
     misc_feature    1243. .2438
                     /note "assembly_fragment"
     misc_feature    2539. .3714
                     /note "assembly_fragment"
     misc_feature    3815. .5559
                     /note "assembly_fragment"
     misc_feature    5660. .7245
                     /note "assembly_fragment"
     misc_feature    7346. .9329
                     /note "assembly_fragment"
     misc_feature    9430. .11234
                     /note "assembly_fragment"
     misc_feature    11335. .12701
                     /note "assembly_fragment"
     misc_feature    12802. .14171
                     /note "assembly_fragment"
     misc_feature    14272. .15608
                     /note "assembly_fragment"
     misc_feature    15709. .17232
                     /note "assembly_fragment"
     misc_feature    17333. .18977
                     /note "assembly_fragment"
     misc_feature    19078. .21304
                     /note "assembly_fragment"
     misc_feature    21405. .25252
                     /note "assembly_fragment"
     misc_feature    25353. .28561
                     /note "assembly_fragment"
     misc_feature    28662. .31363
                     /note "assembly_fragment"
     misc_feature    31464. .35304
                     /note "assembly_fragment"
     misc_feature    35405. .39966
                     /note "assembly_fragment"
     misc_feature    40067. .42541
                     /note "assembly_fragment"
     misc_feature    42642. .46458
                     /note "assembly_fragment"
     misc_feature    46559. .51258
                     /note "assembly_fragment"
     misc_feature    51359. .54202
                     /note "assembly_fragment
                     clone_end:3P
                     vector_side:right"
     misc_feature    54303. .59459
                     /note "assembly_fragment"
     misc_feature    59560. .63206
                     /note "assembly_fragment
                     clone end: ?
                     vector_side:right"
     misc_feature    63307. .67217
                     /note "assembly_fragment"
     misc_feature    67318. .72748
                     /note "assembly_fragment"

 Query Match              52.1%  Score 22.6  DB 1  Length 139594;
 Best Local Similarity    72.5%  Pred. No. 1.
 Matches   29;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       3 ttcgtcaqqcaca-taatqqtcaacqqttqatacatc 41
           ||  || |||||||||| || | ||||  ||| |
Db  119211 TTTAACATTCACGCTAATCGGTTTTAATCGTCAAC 119247

Search completed: March 14, 2001, 15:00:47
Job time: 2382 sec
```